

LAPORAN ANALISIS & HASIL PENGOLAHAN DATA (PYTHON + DATA MINING)

1 Deskripsi Dataset

- Sumber dataset : <https://github.com/wildangunawan/dataset-saham-idx>
- Jumlah record : Menyesuaikan dengan periode waktu dan jumlah emiten dalam dataset (ribuan hingga puluhan ribu baris data).
- Jumlah atribut : $\pm 7-9$ atribut utama.
- Tipe data : Numerik dan Date
- Target/label (jika supervised): Klasifikasi arah harga saham (Naik, Turun, Stabil)
- Permasalahan yang ingin diselesaikan: Memprediksi arah pergerakan harga saham berdasarkan data historis, Menentukan algoritma data mining yang paling optimal untuk klasifikasi pergerakan saham IDX.

“Dataset Saham IDX digunakan untuk memprediksi arah pergerakan harga saham (naik, turun, atau stabil) berdasarkan data historis harga dan volume transaksi.”

2 Persiapan Data & Preprocessing

Jelaskan langkah preprocessing yang Anda lakukan:

- **Data cleaning** (missing value, outlier): Menghapus data duplikat, Menangani missing value pada kolom harga dan volume, Mengurutkan data berdasarkan tanggal.
- **Encoding** data kategorikal: Kode saham (Ticker) \rightarrow LabelEncoder.
- **Scaling / Normalization** (StandardScaler): Menggunakan StandardScaler untuk fitur numerik (Open, High, Low, Close, Volume).
- **Feature engineering**: Membuat fitur baru: Price_Change = Close - Open, Daily_Return = (Close - Open) / Open, Membuat label klasifikasi: Naik \rightarrow Close > Open, Turun \rightarrow Close < Open, Stabil \rightarrow Close \approx Open.
- **Split data**: data training 80% & data testing 20%.

Tahapan	Sebelum	Sesudah
Missing Value	Ada	Ditangani
Fitur Numerik	Belum diskalakan	Sudah distandardisasi
Label	Belum ada	Naik / Turun / Stabil
Data	Mentah	Siap modeling

3 Analisis Statistik & Visualisasi

Sertakan:

- Statistik deskriptif dataset: Rata-rata harga saham menunjukkan fluktuasi signifikan, Volume transaksi bervariasi antar emiten dan waktu.
- Visualisasi pendukung (histogram, boxplot, heatmap): Histogram distribusi harga Close, Boxplot untuk mendeteksi outlier harga saham, Heatmap korelasi antar fitur harga, Line chart tren harga saham per waktu

Harga Open, High, Low, Close memiliki korelasi sangat kuat, Volume transaksi tidak selalu berbanding lurus dengan kenaikan harga, Fluktuasi harga saham menunjukkan pola non-linear.

4 Pemilihan dan Penerapan Algoritma

Tuliskan:

- Nama algoritma: Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM).
- Alasan pemilihan: Dataset bersifat numerik & non-linear, Random Forest cocok untuk data dengan noise, KNN sebagai pembanding sederhana, SVM efektif untuk pola kompleks.

Algoritma	Library Python	Tujuan
K-Nearest Neighbors (KNN)	sklearn.neighbors	Klasifikasi
Random Forest	sklearn.ensemble	Klasifikasi & Feature Importance
Support Vector Machine (SVM)	sklearn.svm	Klasifikasi non-linear

5 Pengujian dan Evaluasi Model

Metode evaluasi: Accuracy, Precision, Recall, F1-Score, Confusion Matrix.

Contoh Tabel Hasil Klasifikasi

Algoritma	Accuracy	Precision	Recall	F1-Score
KNN	0.85	0.84	0.83	0.83
Random Forest	0.91	0.92	0.90	0.91
SVM	0.89	0.88	0.87	0.88

6 Analisis & Interpretasi Hasil

Analisis data mining:

1. Random Forest memberikan performa terbaik dengan akurasi tertinggi.
2. Fitur yang paling berpengaruh:
 1. Close
 2. Open
 3. Daily Return
3. Model tidak menunjukkan indikasi overfitting signifikan.
4. Pola pergerakan saham IDX cenderung non-linear.

“Random Forest mampu memberikan performa terbaik karena dapat menangani kompleksitas dan fluktuasi data saham yang bersifat non-linear.”

7 Kesimpulan & Rekomendasi

Kesimpulan

1. Dataset Saham IDX berhasil digunakan untuk klasifikasi arah pergerakan harga saham.
2. Random Forest merupakan algoritma paling optimal pada penelitian ini.

Rekomendasi

1. Menambahkan data historis yang lebih Panjang.
 2. Melakukan hyperparameter tuning.
 3. Mencoba algoritma deep learning (LSTM) untuk time series.
 4. Mengatasi kemungkinan class imbalance.
-

Lampiran (Opsional)